



Finding Complementary and Reliable Patterns in Displacement Field Time Series of Alpine Glaciers

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier,
Emmanuel Trouvé, Jean-Louis Mugnier

► To cite this version:

Tuan Nguyen, Nicolas Méger, Christophe Rigotti, Catherine Pothier, Emmanuel Trouvé, et al.. Finding Complementary and Reliable Patterns in Displacement Field Time Series of Alpine Glaciers. IGARSS 2018 - IEEE International Geoscience and Remote Sensing Symposium, Jul 2018, Valencia, Spain. pp.4213-4216. hal-01868749

HAL Id: hal-01868749

<https://hal.science/hal-01868749>

Submitted on 5 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FINDING COMPLEMENTARY AND RELIABLE PATTERNS IN DISPLACEMENT FIELD TIME SERIES OF ALPINE GLACIERS

T. Nguyen^{1*}, N. Méger¹, C. Rigotti², C. Pothier³, E. Trouvé¹ and J.-L. Mugnier⁴

¹Univ. Savoie Mont Blanc, LISTIC, F-74000 Annecy, France

²Univ Lyon, INSA-Lyon, CNRS, INRIA, LIRIS, UMR5205, F-69621, France

³Univ Lyon, INSA-Lyon, CNRS, LIRIS, UMR5205, F-69621, France

⁴Univ. Savoie Mont Blanc, CNRS, ISTerre, F-73376 Le Bourget-du-Lac, France

ABSTRACT

Dynamic systems such as glaciers can be studied using *Displacement Field Time Series (DFTS)*, often derived from *Satellite Image Time Series*. Even if data mining patterns expressing interesting displacement evolutions can be extracted from DFTS, confidence measures coming along with these series have to be considered to focus on reliable evolutions. This paper introduces a new approach for selecting displacement evolutions that are reliable, informative and complementary. Reported experiments exhibit consistent displacement evolutions of Alpine glaciers and complete the current knowledge of the area.

I. INTRODUCTION

Displacement Field Time Series (DFTS), usually calculated from *Satellite Image Time Series (SITS)*, can be analyzed using sequential pattern mining techniques, allowing to discover meaningful displacement evolutions over time and space [1]. Besides handling space and time, DFTS analysis techniques also have to consider the confidence measures associated with such series. These measures can be, for example, computed by evaluating the spatial and/or temporal coherence of the displacement vectors (e.g., [2]). One of the main difficulties is that they cannot be handled as a single global quality measure of the DFTS, since, as for displacement data themselves, the confidence can change over space and time. A pattern-based method has thus been recently introduced in [3] to extract displacement evolutions having a confidence exceeding a given threshold. The approach has two limitations: it does not handle the confidence during the ranking of the patterns and delivers a set of patterns that may contain many redundant evolutions.

The main contributions of this paper are the following: (1) the definition of the information gain obtained from the knowledge of the occurrences of a pattern according to the DFTS confidence measures, (2) a heuristic for building a set of reliable and informative patterns that complement each other, and (3) experiments on Alpine glaciers showing that the method can help end-users to complete the current knowledge of the area.

II. DFTS AND RELIABLE GFS-PATTERNS

We consider a *symbolic displacement field time series* containing for each location (x, y) a *displacement evolution sequence* $seq(x, y) = \langle (t_1, \alpha_1, \rho_1), \dots, (t_n, \alpha_n, \rho_n) \rangle$, where α_i and $\rho_i \in [0, 1]$ (with

$i \in \{1, \dots, n\}$) are respectively the symbol encoding the displacement magnitude and its confidence value.

As explained in [1], the *Grouped Frequent Sequential Patterns (GFS-patterns)* originally designed to analyze SITS [4] can be adapted to process such symbolic DFTS. A *sequential pattern* β is a pattern of the form $\beta_1 \rightarrow \beta_2 \rightarrow \dots \rightarrow \beta_m$ where β_1, \dots, β_m are m symbols. Let $seq(x, y) = \langle (t_1, \alpha_1, \rho_1), \dots, (t_n, \alpha_n, \rho_n) \rangle$ be a displacement evolution sequence. Location (x, y) is covered by β if there exist $i_1 < i_2 < \dots < i_m$ so that $\beta_1 = \alpha_{i_1}, \beta_2 = \alpha_{i_2}, \dots, \beta_m = \alpha_{i_m}$. In this case, $o = \langle (t_{i_1}, \rho_{i_1}), (t_{i_2}, \rho_{i_2}), \dots, (t_{i_m}, \rho_{i_m}) \rangle$ is called an *occurrence* of β in $seq(x, y)$. In order to consider core displacement evolutions, *minimal occurrences*, as defined in [5], are focused on. An occurrence o of pattern β is called *minimal occurrence* if β does not occur in any proper subinterval of $[t_{i_1}, t_{i_m}]$. The GFS-patterns [4] are the sequential patterns that are *frequent* and *grouped*. The term *frequent* means that a pattern covers at least σ locations (x, y) in the studied area (i.e., it covers at least a minimum surface). A pattern β is said to occur in a *grouped* way if, on average, a location covered by β is surrounded by at least κ other locations in its 8-neighborhood that are also covered by β .

In order to deal with confidence measures and extract meaningful patterns, a notion of pattern reliability has been introduced in [3]. A pattern is said to be *reliable* if it is built on displacement values whose corresponding confidence measures reach, on average, a minimum level of confidence noted γ . The reader is referred to [3] for formal definitions.

III. PATTERN SET SELECTION

In this section, we propose an original method for finding a set of informative and complementary patterns that takes into account the confidence measures. Let us consider a symbolic DFTS containing $|T|$ acquisition dates and covering Z locations. Its information content can be expressed as the entropy of a set of $|T| \times Z$ random variables, noted $\mathcal{V} = \{X_1, X_2, \dots, X_{|T| \times Z}\}$. Suppose that we only have a partial knowledge of the series. The key intuition of the method is that if we are given the occurrences of a pattern β , then this knowledge can provide additional information about the distributions of variables in \mathcal{V} and reduce the uncertainty we have about the series. The larger this reduction is, the more informative the pattern is. In the following, this gain with respect to the current partial information we have about the series is noted $\Delta(\beta)$.

Finding an optimal set of patterns with respect to an entropy criterion is in general NP-hard [6]. Inspired by the *SeqKrimp* algorithm [6], designed to find sets of patterns that compress datasets, we use a greedy suboptimal algorithm, given as Algorithm 1, that chooses the most informative patterns in an iterative way.

In Algorithm 1, line 2, the distributions of all variables in \mathcal{V} are initialized to the uniform distribution over domain I , since, at the beginning, the only known information is that the discretization has been made using equal frequency bucketing. This state corresponds to a maximum of entropy.

For a pattern β , the gain $\Delta(\beta)$, used line 4, is defined as follows. Let $\Delta_{occ}(o)$ denote the gain for an occurrence o of β . It is based on the occurrence quality and is defined hereafter. In a sequence $seq(x, y)$ covered by β , we consider that it is the *best* occurrence of β that will be output to the user. This best occurrence is simply defined as the one leading to the highest gain. Thus, for the user, the gain coming from β in $seq(x, y)$ is $\Delta(seq(x, y), \beta) = \max_{o \in \mathcal{O}} \{\Delta_{occ}(o)\}$, where \mathcal{O} is the set of occurrences of β in $seq(x, y)$.

Then, the measure of the gain $\Delta(\beta)$, which assesses the interest of this pattern over all sequences it covers, is defined as:

$$\Delta(\beta) = \frac{\sum_{(x,y) \in cov(\beta)} \Delta(seq(x, y), \beta)}{|cov(\beta)|}$$

Algorithm 1 Selection of k informative and complementary patterns

Input: k the number of patterns to select, P the set of reliable GFS-patterns, and their occurrences extracted from the symbolic DFTS

Output: Φ a set of informative and complementary patterns

- 1: $\Phi \leftarrow \emptyset$
 - 2: $\forall X \in \mathcal{V}$, set distribution of X to the uniform distribution
 - 3: **while** ($|\Phi| < k$ and $|P| > 0$) **do**
 - 4: $\beta^* \leftarrow \operatorname{argmax}_{\beta \in P}(\Delta(\beta))$, i.e., the pattern maximizing gain w.r.t. the current distributions known for the variables in \mathcal{V}
 - 5: $\Phi \leftarrow \Phi \cup \{\beta^*\}$
 - 6: $P \leftarrow P \setminus \{\beta^*\}$
 - 7: $\forall X \in \mathcal{V}$, update distribution of X
 - 8: **end while**
 - 9: **return** Φ
-

It should be noticed that the measure is divided by $|\operatorname{cov}(\beta)|$, with $\operatorname{cov}(\beta)$ the set of locations that are covered by β . The intuition for this normalization is that a pattern covering many locations with occurrences of poor quality is likely to be less desirable than a pattern covering a smaller area but relying on data points having higher confidences.

To define $\Delta_{occ}(o)$, the gain for an occurrence o of pattern β , we interpret the DFTS as a probabilistic database. In this context, an element (t, α, ρ) in a sequence of the DTFS denotes that symbol α appears at timestamp t with probability ¹ ρ . An occurrence o then simply reveals for some of the data points the symbols and their associated probabilities. The information gain provided by o is the advantage we obtain by knowing the symbols and their probabilities (as given by o), when compared to someone that has less knowledge about the distributions of the symbols over the DFTS.

For example, let us consider a sequence seq in a DFTS over a domain I containing three symbols: 1, 2 and 3. Let us suppose that seq contains five elements from timestamp t_1 to t_5 , each timestamp being associated to a random variable X_{t_1} to X_{t_5} . At the beginning, for instance for variable X_{t_3} , since no other information is available, a distribution p having the highest entropy is retained for this variable. This distribution p is simply the uniform distribution over $\{1, 2, 3\}$.

Let us now suppose that for a pattern $\beta = 1 \rightarrow 3 \rightarrow 2$ we are given the occurrence $o = \langle (t_1, 0.9), (t_3, 0.9), (t_5, 0.7) \rangle$. This adds to our knowledge of the true distribution of X_{t_3} , since it means that the second element of the pattern, the symbol 3, occurs at timestamp t_3 with probability 0.9. The updated distribution q for X_{t_3} is then such that $Pr(X_{t_3} = 3) = 0.9$. Since we have no information about the other symbols, their probabilities are chosen so as to obtain the minimum information gain, i.e., their values must lead to the highest entropy. We thus obtain $Pr(X_{t_3} = 1) = Pr(X_{t_3} = 2) = 0.1/2 = 0.05$. Using the occurrence o , similar refinements can be made for the distributions of the variables X_{t_1} and X_{t_5} .

More complex constraints on the distributions can also be implied, depending on the kind of occurrences. A more general approach is thus adopted and briefly introduced as follows. Consider one of the random variable X_t and its current known distribution p . To determine $\Delta_{occ}(o)$, the gain for an occurrence o , we derive the set η of constraints implied by o on the real (but unknown) distribution of X_t . The problem is then to find a distribution q for X_t that is the closest to p (i.e., adding the smallest

¹Depending on the meaning of confidence ρ , the probability can be ρ itself or a value derived from ρ .

amount of information) but satisfying the set of constraints η (that must hold for the real distribution). To this aim we use the *Kullback-Leibler* divergence, defined as follows:

$$D(q||p) = \sum_{\alpha \in I} q(\alpha) \log_2 \left(\frac{q(\alpha)}{p(\alpha)} \right)$$

where $p(\alpha)$ (resp. $q(\alpha)$) is $Pr(X_t = \alpha)$ in distribution p (resp. q). $D(q||p)$ quantifies *the information gained* if the distribution p is replaced by the real distribution q [7]. The distribution q we are looking for is then a distribution satisfying the constraints η and minimizing $D(q||p)$, so as to be as close as possible to p (from an information content perspective). It is obtained by a constraint-based optimization scheme.

The value of $\Delta_{occ}(o)$ is then the sum of the minimized Kullback-Leibler divergences, over all the pairs (p, q) for the random variables in \mathcal{V} that are involved in the constraints derived from occurrence o .

As defined previously, $\Delta_{occ}(o)$ for all occurrences of β is used to determine $\Delta(\beta)$ that is maximized (line 4 in Algorithm 1). Then, the current sets of patterns Φ and P are updated (lines 5 and 6), and finally, line 7, the distributions of the random variables are modified according to the occurrences of the selected pattern β^* . The computation of the new distributions follows the same procedure as the determination of q from p given above, by minimization of the Kullback-Leibler divergence under the constraints implied by the occurrences of β^* .

IV. EXPERIMENTS

The proposed approach has been applied to explore displacement evolutions of Alpine glaciers in the Mont Blanc massif. The DFTS is built from 26 TerraSAR-X images and is composed of two distinct periods, one in 2009 from May 31 to October 21 and another in 2011 from May 5 to September 25. Each period contains 13 images, with one image every 11 days. *Synthetic Aperture Radar* (SAR) images contain amplitude and phase information. The former is used to compute displacement fields, with the amplitude cross-correlation method implemented in the *EFIDIR Tools*². 25 displacement fields, expressed in *m/day*, are obtained. Following the works of [2], the confidence measure is intended to reflect the *temporal directional coherence* of the 2D displacement vectors. For each sequence $seq(x, y)$, and for each timestamp t , the confidence measure of the displacement ρ_{disp} is defined as:

$$\rho_{disp}(x, y, t) = \max \left(\cos(\widehat{\vec{v}_{x,y,t}, \vec{u}_{x,y}}), 0 \right)$$

where $\vec{v}_{x,y,t}$ is the displacement vector and $\vec{u}_{x,y} = \sum_{t=1}^n \frac{\vec{v}_{x,y,t}}{|\vec{v}_{x,y,t}|}$. The larger is the angle, the smaller is the confidence.

In order to get robust estimates, the original size of the 25 displacement fields containing the values $\vec{v}(x, y, t)$ is reduced using a low-pass filtering and sub-sampling. Let $\Omega_{i,j,t}$ be the list of the values of $\|\vec{v}(x, y, t)\|$ contained in a 3x3 tiling window $w_{i,j,t}$, and $\Omega_{i,j}$ be the concatenation of the lists $\Omega_{i,j,1}, \Omega_{i,j,2}, \dots, \Omega_{i,j,25}$ (i.e., all values over time in windows of indices (i, j)). Let $MAD_{i,j}$ denotes the Median Absolute Deviation at location (i, j) . Then, the values describing the displacement fields are the median differential speeds available for each timestamp t and obtained for each window by:

$$mds(i, j, t) = \frac{\text{median}(\Omega_{i,j,t}) - \text{median}(\Omega_{i,j})}{MAD_{i,j}}$$

²<http://efidir.poleterresolide.fr/index.php/efidir-tools>

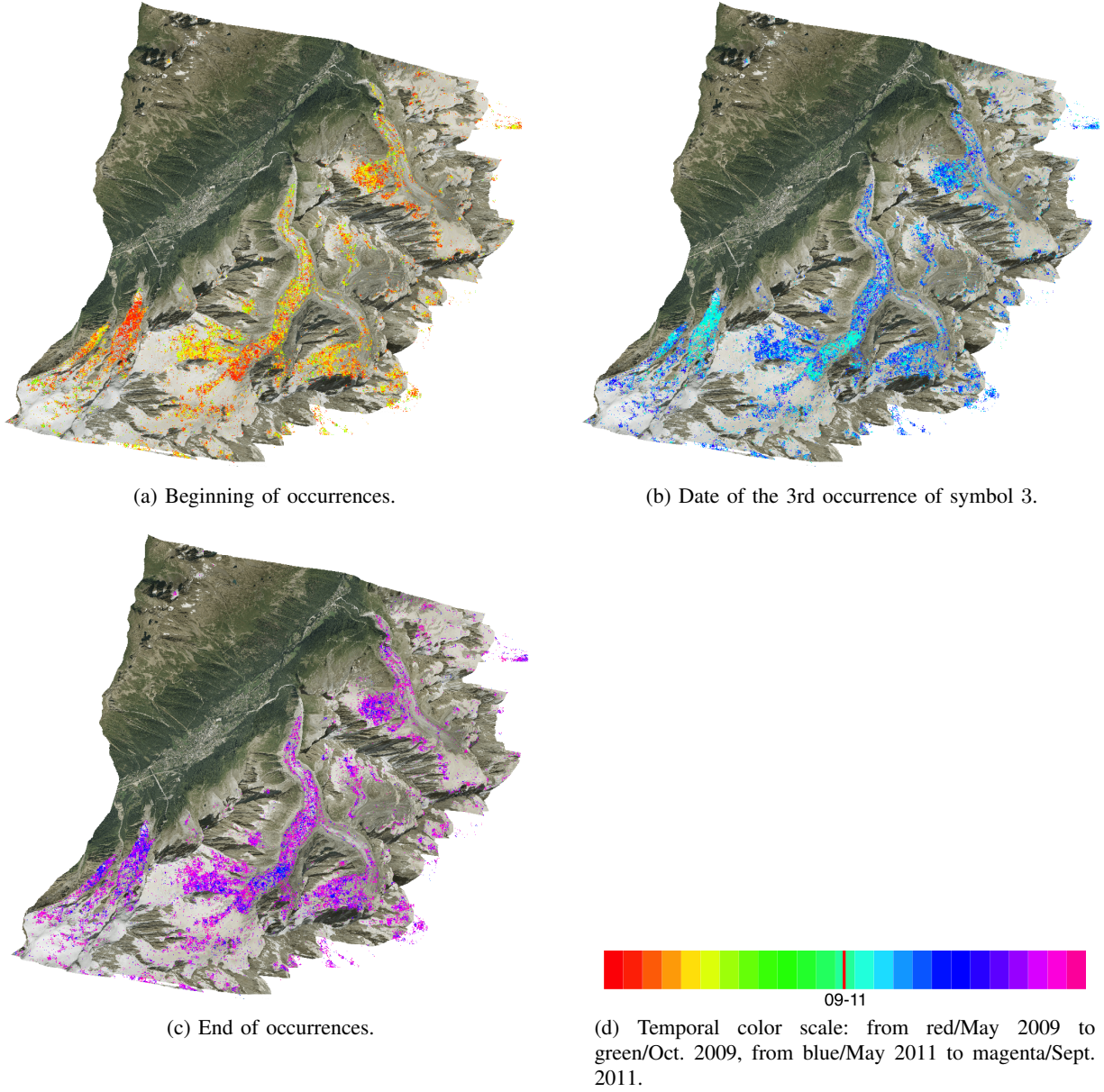


Fig. 1: Maps of pattern $3 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2$ showing in synthetic colors where and when the pattern occurs according to the temporal color scale. The background is a RGB composite projected in radar geometry.

The symbolic series is then built from the values of $mds(i, j, t)$ by quantization, using an equal frequency bucketing and symbols 1, 2 and 3 (denoting respectively low, medium and high values). For the confidence, the measure $\rho(i, j, t)$ is obtained from the median of the confidence ρ_{disp} over the window $w_{i,j,t}$. According to the method given in [3], the extraction parameters of the reliable GFS-patterns were set to $\sigma = 440500$, $\kappa = 5$ and $\gamma = 0.22$.

Then, the selection of the top 20 informative and complementary patterns was performed and different evolutions over the velocity sequences were obtained. One is $3 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2$ which exhibits two slowdowns over time: one from level 3 to level 1 and then from level 3

to level 2. The locations where this pattern occurs (in space) are depicted by colored dots in Figure 1. Using the temporal color scale shown in Figure 1d, the timestamps (from 1 to 25) corresponding to the beginning (resp. to the end) of the occurrences are given in Figure 1a (resp. Figure 1c). The first slowdown captured by the pattern, i.e., the $3 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 1$ part, begins in early summer 2009 (timestamps shown in Figure 1a). The second slowdown, i.e., the $\rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2$ part, starts in 2011 according to the timestamps of this symbol 3 depicted in Figure 1b. Such slowdowns are known glaciological processes for temperate glaciers (e.g., [8]). In the Mont Blanc massif, they have been reported by [9] along three transects for the year 2009. Here, the pattern shows that the slowdown has been repeated in 2011, and beyond a transect-based analysis, it underlines the 2D spatial extent of the phenomenon.

The whole process takes about 36 hours to obtain the 20 best ranked evolution patterns using an Intel Xeon 3.5 GHz (16 cores) running Linux (Ubuntu).

V. CONCLUSION

The quality of displacement field time series computed from SAR images can be reflected by the confidence measure associated to each data point. In this paper, we show that this measure can be exploited to select informative and complementary displacement evolution patterns. To this end, each data point is assumed to be a random variable whose discrete probability distribution is refined by combining the information about the confidence of pattern occurrences. The corresponding information gain is quantified using the Kullback-Leibler divergence. Finally, a greedy algorithm is proposed so as to select informative patterns that complement each other. Experiments on Alpine glaciers show that meaningful patterns can be captured.

ACKNOWLEDGEMENTS

Funding for this project was provided by a grant from la Région Auvergne-Rhône-Alpes (Tuan Nguyen's grant). The work was also supported by a University Savoie Mont Blanc AAP Montagne grant - project BG *Big data pour la surveillance des Glaciers*. Catherine Pothier and Christophe Rigotti are members of LabEx IMU (ANR-10-LABX-0088) that provided complementary support. The authors would like to thank the German Aerospace Center (DLR) for providing the TerraSAR-X time series used in the experiments under Project MTH0232.

VI. REFERENCES

- [1] Y. Pericault et al., "A swap randomization approach for mining motion field time series over the Argentiere glacier," in *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 2015.
- [2] D. Scherler et al., "Glacier-surface velocities in alpine terrain from optical satellite imagery - Accuracy improvement and quality assessment," *Remote Sensing of Environment*, 2008.
- [3] T. Nguyen et al., "Handling coherence measures of displacement field time series: Application to Greenland ice sheet glaciers," in *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, 2017.
- [4] A. Julea et al., "Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns," *IEEE Transactions on Geoscience and Remote Sensing*, 2011.
- [5] H. Mannila et al., "Discovery of frequent episodes in event sequences," *Data mining and knowledge discovery*, 1997.
- [6] H. T. Lam et al., "Mining Compressing Sequential Patterns," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2014.
- [7] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, 1961.

- [8] C. Vincent et al., “Sliding velocity fluctuations and subglacial hydrology over the last two decades on Argentière glacier, Mont Blanc area,” *Journal of Glaciology*, 2016.
- [9] R. Fallourd, *Suivi des glaciers alpins par combinaison d’informations hétérogènes: images SAR Haute Résolution et mesures terrain*, Ph.D. thesis, Université de Grenoble, 2012.